CLAIMS

That which is claimed:

1. A method, comprising:

selecting a plurality of items, each item having an entry in an inverted index, each item entry comprising a listing of articles that the item appears in;

determining at least a first item entry and a second item entry for compression; and

compressing the second item entry into the first item entry resulting in a compressed first item entry.

2. The method of claim 1, wherein determining the first item entry and the second item entry for compression comprises:

determining a cost-benefit ratio for compressing the second item entry into the first item entry; and

comparing the cost-benefit ratio with an acceptable value to determine if the cost-benefit ratio is acceptable.

3. The method of claim 2, wherein the item entry further comprises an item value for each article that the item appears in.

4. The method of claim 3, wherein the item value is a representation of the strength of the item in the article.

5. The method of claim 3, wherein the item value is a representation of whether the item appears in the article.

6. The method of claim 3, wherein a cost for the cost-benefit ratio comprises a representation of the loss in precision that can be caused by compressing the entries or the additional processing time that can be required when utilizing a compressed entry.

7. The method of claim 6, wherein determining a cost for the cost-benefit ratio comprises determining how much the first item entry and the second item entry have to change when the second item entry is compressed into the first item entry.

8. The method of claim 2, wherein a benefit for the cost-benefit ratio is a representation of the amount of memory saved if the first and second item entries were compressed.

9. The method of claim 1, wherein the items comprise words, concepts or images.

10. The method of claim 2, wherein the acceptable value is predetermined.

11. A method, comprising:

selecting a plurality of items, each item having an entry in an inverted index, each item entry comprising a listing of articles that the item appears in and an item value for each article that the item appears in;

determining at least a first item entry and a second item entry for compression by determining a cost-benefit ratio for compressing the second item entry into the first item entry and comparing the cost-benefit ratio with an acceptable value to determine if the cost-benefit ration is acceptable; and

if the cost-benefit ratio is acceptable, compressing the second item entry into the first item entry resulting in a compressed first item entry.

12. The method of claim 11, wherein determining a cost for the cost-benefit ratio comprises determining how much the first item entry and the second item entry have to change when the second item entry is compressed into the first item entry..

13. The method of claim 11, wherein a benefit for the cost-benefit ratio is a representation of the amount of memory saved if the first and second item entries were compressed.

14. The method of claim 11, wherein the acceptable value is predetermined.

15. The method of claim 11, wherein the items comprise words, concepts or images.

16. A method, comprising:

selecting a plurality of items, each item having an entry in an inverted index, each item entry comprising a listing of articles that the item appears in;

determining a plurality of item entries for compression; and

compressing the item entries into a compressed item entry.

17. The method of claim 16, wherein the plurality of item entries comprises three or more item entries.

18. A computer-readable medium containing program code, comprising:

program code for selecting a plurality of items, each item having an entry in an inverted index, each item entry comprising a listing of articles that the item appears in;

program code for determining at least a first item entry and a second item entry for compression; and

program code for compressing the second item entry into the first item entry resulting in a compressed first item entry.

19. The computer-readable medium of claim 18, wherein determining the first item entry and the second item entry for compression comprises:

program code for determining a cost-benefit ratio for compressing the second item entry into the first item entry; and

program code for comparing the cost-benefit ratio with an acceptable value to determine if the cost-benefit ratio is acceptable.

20. The computer-readable medium of claim 17, wherein the item entry further comprises an item value for each article that the item appears in.

21. The computer-readable medium of claim 20, wherein the item value is a representation of the strength of the item in the article.

22. The computer-readable medium of claim 19, wherein the item value is a representation of whether the item appears in the article.

23. The computer-readable medium of claim 22, wherein a cost for the cost-benefit ratio comprises a representation of the loss in precision that can be caused by compressing the entries or the additional processing time that can be required when utilizing a compressed entry.

24. The computer-readable medium of claim 20, wherein determining a cost for the cost-benefit ratio comprises program code for determining how much a first item entry and a second item entry have to change when the second item entry is compressed into the first item entry.

25. The computer-readable medium of claim 19, wherein a benefit for the cost-benefit ratio is a representation of the amount of memory saved if the first and second item entries were compressed.

26. The computer-readable medium of claim 18, wherein the items comprise words, concepts or images.

27. The computer-readable medium of claim 19, wherein the acceptable value is predetermined.

28. A computer-readable medium containing program code, comprising:

program code for selecting a plurality of items, each item having an entry in an inverted index, each item entry comprising a listing of articles that the item appears in;

program code for determining a plurality of item entries for compression; and

program code for compressing the item entries into a compressed item entry.

29. The computer-readable medium of claim 28, wherein the plurality of item entries comprises three or more item entries.